ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

INFORMATION RETRIEVAL USING : HIERARCHICAL CLUSTERING ALGORITHM

D.Saravanan*
Associate Professor, Dept of Operations & IT , IFHE University, IBS Hyderabad 501203 Telangana.
Email: Sa_roin@yahoo.com

Received on: 20.10.2016 *Accepted on: 25.11.2016*

Abstract

In the past decade, more and more information has been published in computer readable formats. The greatest challenge of the WWW is that the more information available about a given topic, the more difficult it is to locate accurate and relevant information. The query will be image or text depended upon the query result output system that returned a particular video or image based on that query, for the effective retrieval the information is clustered effectively. In this paper, we discuss one of the clustering mechanism, discover natural clusters that many existing clustering algorithms fail to find Experimental results show confirms that the proposed method is most effective. This paper focus on the problem of information retrieval. This conducted in two phases client ortraning phase and reterival phase.

Keywords: Data mining; Hierarchical Clustering; Chemeleon; shots; Hierarchical clustering.

1. Introduction

Recently data mining of video databases is widely used in various applications **s**uch as geographical information system (GIS), crime prevention, web searching, cultural heritage, advertising, news broadcasting, video, education and training and military. These applications have an enormous collection of images in the related video databases and also it could be mined to determine new and valuable knowledge process[1]. Such a knowledge discovery process is automated by the data mining of video database, which help the users to search and retrieve the relevant video effectively and to enhance the better ways of entertainment. In this situation very urgent requirement needed to retrieve the contents most effectively. However, there exists valuable information behind the video data; the increasing data volume creates complexity to humans in extracting them without sufficient tools. Hierarchical clustering is divided into two types, namely agglomerative and divisive hierarchical clustering, depending on whether the breakdown is formed in a bottom-up or top-down fashion[2].

Quality of the cluster is considered by the distance between the each pair of group or subgroup clusters, they way in which image registration are performed[3]. Agglomerative deals clustering in a single search and cluster all the data or objects in a single object. It is a step by step approach. Series of merging operations done as a result got a single cluster. But in case of divisive clustering is opposite to the early process. Find out the data sets at a time and divide the data set of successive divisions until reach singleton cluster or the stage the data to be non- separable. For video files the agglomerative clustering process offers a good clustering result.

2. Classifications of Video

Classification is nothing but the method of categorizing or classifying the class labels to a pattern set under the supervision. Decision boundaries are usually created to distinguish the patterns among the various classes. Initially the datasets are divided into a number of segments, and then the classifiers are trained.

The research work was proposed to enhance the semantic video classification and indexing in a particular domain. The techniques to classify the various types of videos, which make use of the result of video summarization technique that form the list of key frames

2.1 Clustering of Video Data

Generally the clustering is a useful technique to obtain some information from the given dataset. It maps the similar data items into clusters. The clusters are formed by grouping the data items that are based on probability density models and similarity metrics.

Clustering consists of dividing the data into homogenous groups or else granules, which depend on the same objective function which maximize the inter cluster distance. Consequently the video clustering is differing from the traditional clustering algorithms.

However the videos are in the unstructured format. Thus preprocessing the video data is necessary to obtain the structured format of the video using the computer vision techniques or image processing.

One more difference in the video clustering is the time factor. It must be taken as a parameter when processing the video data. Conventional clustering algorithms can be classified into two major types such as hierarchical and partitioned clusters.

All clusters are determined only one time by the hierarchical clustering technique. It may be in the bottom up (agglomerative) or top down (divisive). Agglomerative algorithm considers every element as the separate cluster and combines them into large clusters.

Whereas the divisive algorithms start process with the entire data set and continue to partition it into smaller clusters. Co-clustering or bi-clustering and two-way clustering are the clustering methods that not only cluster the objects, but also feature of the objects that are clustered. For instance, consider if the video data is represented in the matrix, the columns and rows are clustered concurrently.

2.3 Video Clustering

Existing clustering technique fail to retrieve the needed information based on the user's input image query. Number of clustering and retrieval technique proposed in this concern. To retrieve the effective content the input data sets are grouped, it avoids the searching time.

Clustering the video data's is not an easy task. Multimedia content consists of the complex type of data's among those creating the grouping is one of the challenge job[4,5] for doing this process effectively we need a video clustering technique.

This technique differs from the normal clustering techniques. Because video clustering involves the unstructured data set. Due to this complex data set grouping take a long time. In this proposed method we use hierarchical clustering technique.

It allows the user to represent the data set in the hierarchical fashion, and also allow backtracking. Due to this user can perform the operations very effectively. Increasing the quality and the quality of the video files, content based techniques most helpful[6].

3. Existing System:

Birch Method of forming Cluster : Employs tree structure and tunes the quality of sub clusters.

Method of Implementation: BIRCH uses a Clustering Feature (CF) and Clustering Features tree (CF-Tree).

Applicable for: Multi-dimensional metric data objects.

Improvements Required: Process only numeric data and susceptible to the position of the data record

Cure:

Method of forming Cluster: Draws well dispersed objects from the cluster and then moves them towards the cluster center by a certain fraction

Method of Implementation: Cure uses random sampling methods and partitioning approach.

Applicable for: Metric space data set.

Improvements Required: CURE failed to identify the exact clusters in a given database.

3.1 Advantage of Proposed System:

1. Being applicable to all types of data.
2. Two-phase approach
3. It is possible to use other algorithms instead of k-nearest neighbor graph
4. It performs any data set.
5. Chameleon finds initial sub cluster and combining sub clusters by dynamic framework model.This model shown in the Figure 1.

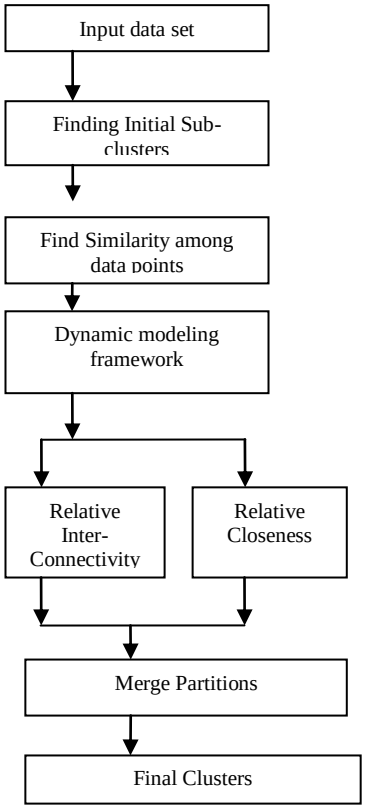


Figure 1. Proposed Clustering Model.

4. Experimental Setup:

Chameleon algorithm is implemented in two different phases. In the first phase, the data objects are grouped into sub clusters by dynamic modeling approach. In the second phase, to obtain good quality cluster, a dynamic framework model is implemented on the data objects to combine the sub clusters in a hierarchical approach. The model can be executed in two separate methods.

First Method**:** The values of relative closeness and relative inter-connectivity between two clusters are examined such that they exceed a user – mentioned threshold value. Second Method: A pair of clusters that maximize the following function is chosen in this method. After segmenting the image, Grey value represents the value of the difference between the two adjacent image grey values. By assuming a threshold for the grey value the duplicate images are

found out[7] Using the file handling method the duplicates files are eliminated. After eliminating duplication the grouping frames are taken as input for proposed technique. In each frame the histogram value is calculated in the g the following manner. This value is stored for retrieval of user input query

4.1. Forming Sub Clusters: In the experimental design first we Finding Initial Sub-clusters It can get the input dataset and apply the spare graph.

4.2. Finding the nearest: After performing the sub clusters perform the similarity between the element available in the data sets.

4.3. Merging sub clusters: After performing the above task we perform the merging operation by using the CHEMELON cluster method.. It can perform Relative Inter-Connectivity and Relative Closeness. Finally, they are combining by Merge Partitions.

The examperimental outcomes are shown in the Figure 2-9 and duplication elimatian and number of cluster formations and results are shown in the Table 1-4.

Pseudo code for merging sub cluster:

```
imgheight = DisplayBM.Height
x1 = imgwidth1 / 2
y1 = imgheight1 / 2
For y = 0 To imgheight - 1
For x = 0 To imgwidth - 1
colorpixel = DisplayBM.GetPixel(x, y)
str = Format(x, "000" & " ")
str1 = Format(y, "000" & " ")
str2 = colorpixel.R
str3 = colorpixel.G
str4 = colorpixel.B
For y1 = 0 To imgheight1 - 1
For x1 = 0 To imgwidth1 - 1
If x1 = x And y1 = y Then
colorpixel = DisplayOBM.GetPixel(x1, y1)
```

```
stro = Format(x1, "000" & " ")

stro1 = Format(y1, "000" & " ")

stro2 = colorpixel.R

stro3 = colorpixel.G

stro4 = colorpixel.B

If stro = str And stro1 = str1 Then

StatusBar1.Text = "Comparing Values...."

If str2 = stro2 And str3 = stro3 And str4 = stro4 Then

cnt = cnt + 1

Me.Refresh()

lblmsg.Visible = True

lblmsg.Text = cnt

Else

GoTo 3

End If

If cnt >= val Then

storedb()

GoTo 2

End If

End If

End If
```

4.4 These Results are extracted by the following Process

Step1: First, the image is given as the input to the from the camera.

Step2: Initially, this image is a raw image where it contains noise.

Step3: Then, the Features like Texture, Color and Shape are extracted by the RGB values

Step4: These features values and database values are matched. If it matched there will be fast retrieval of the image is done. The content of the image is also retrieved

5. Conclusion:

In this paper, we have presented a hierarchical clustering algorithm[8] called CHAMELEON. It takes the know similarity between the data points makes the cluster effectively. The performance of the existing clustering algorithm works on few video files only. But according to the experimental outcome the proposed technique shows improved outcomes. It reduces the time complexity and produced increasing amount of clusters.

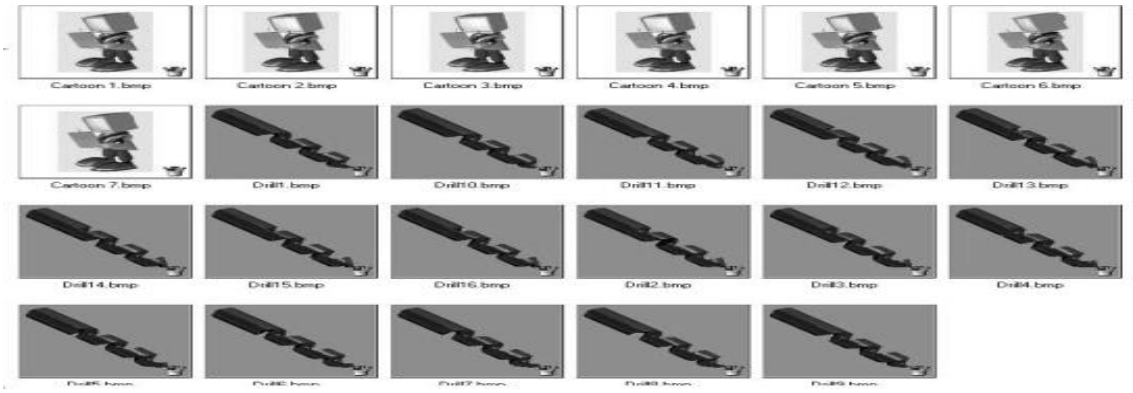
6. Experimental Outcome

6.1 Clustering

The following sample videos are converted into a number of frames for clustering performance comparison of existing techniques with proposed technique.

Frames

1. Cartoon



2. News

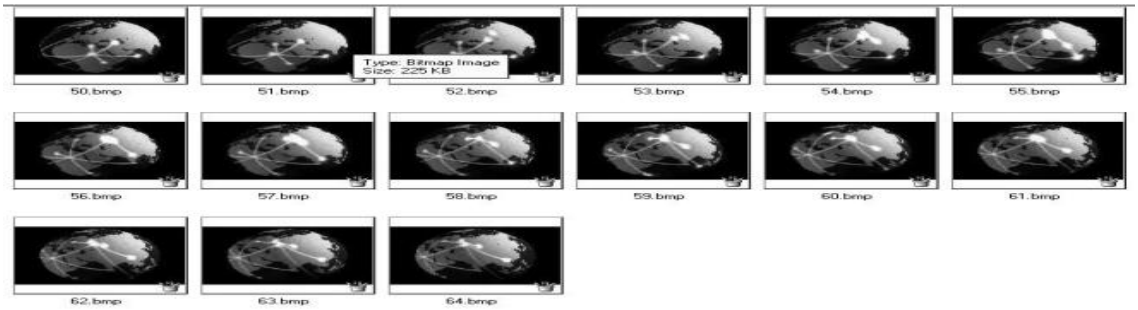


Fig 2. Sample Frame(s) of video files.

Fig 3 .Cluster Formation.

Table-1. Duplicate elimination using Chameleon.

S.no	Video	FrameNo	No of Cluster	Sec
1	1	1	4	60
2	1	2	4	120
3	1	3	5	60
4	1	4	4	120
5	1	5	4	60
6	1	6	6	120
7	1	7	4	60
8	1	8	4	120
9	1	9	6	60
10	1	10	4	60
11	1	11	4	60
12	1	11	6	120
13	1	12	4	120
14	1	13	4	120
15	1	14	6	60
16	2	15	6	180
17	2	1	4	120
18	2	2	4	180
19	2	3	3	180
20	2	4	4	120
21	2	5	5	60
22	2	6	5	120
23	2	7	5	180

fname	value
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \23.bmp	46244344
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \24.bmp	46136769
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \25.bmp	45988813
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \26.bmp	46000717
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \27.bmp	45957601
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \28.bmp	45939020
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \29.bmp	45974799
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \30.bmp	46012732
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \31.bmp	46051725
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \32.bmp	46055509
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \33.bmp	46043439
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \34.bmp	46065575
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \35.bmp	46025120
D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \36.bmp	46031493

Fig 4. Cluster Formation for the input video file.

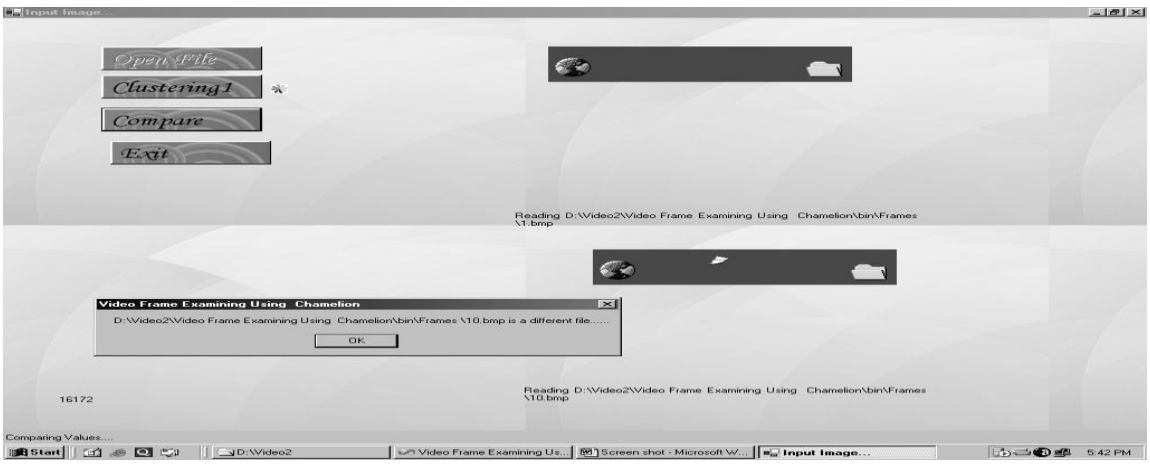


Fig 5. Frame Comparison process.

Table-2 .Results Chameleon Cluster for various video files.

Video name	Number of Input frames	Number of output frames	Duplicate frames removed
Cartoon	7	5	2
Graphics	16	10	6
Meeting	15	14	1
Globe	15	13	2
Song	15	14	1

Table-3. Result of Chameleon.

Frame	Cluster	Mins	Sec
1	6544	0	6
2	6530	0	11
3	6689	0	17
4	6719	0	23
5	6901	0	28

Blue – cartoon

Red – graphics

Rosy brown- movie

Brown – news

Cyan – natural

Table-4. Result of Duplicate Frame Elimination.

Video name	Number of Input frames	Number of output frames	Duplicate frames removed
Cartoon	7	6	1
Graphics	16	10	6
Meeting	15	14	1
Natural Scene	15	13	2
Song	15	14	1

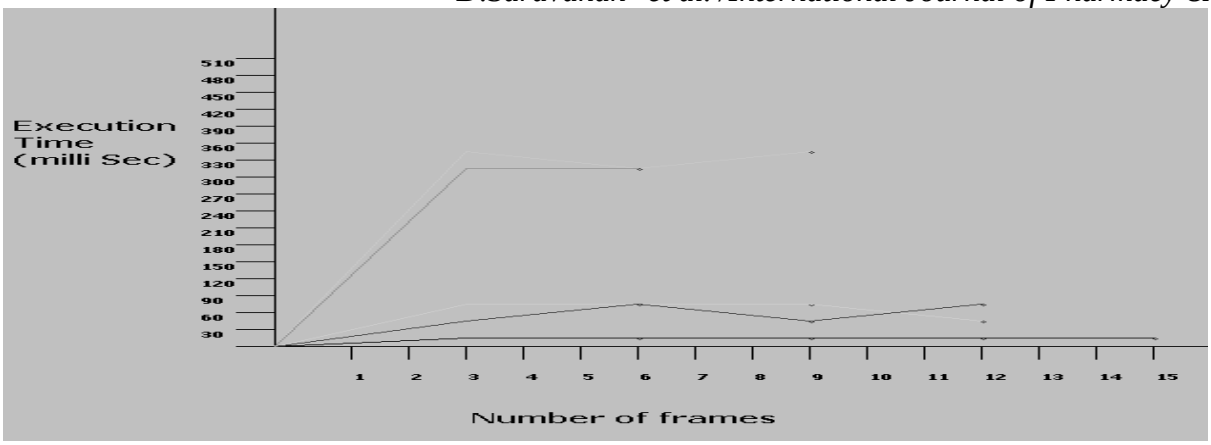


Fig 6. Performance graph of Chameleon.

	fname	value
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \23.bmp	46244344
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \24.bmp	46136769
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \25.bmp	45988813
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \26.bmp	46000717
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \27.bmp	45957601
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \28.bmp	45939020
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \29.bmp	45974799
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \30.bmp	46012732
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \31.bmp	46051725
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \32.bmp	46055509
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \33.bmp	46043439
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \34.bmp	46065575
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \35.bmp	46025120
	D:\Video2\Video Frame Examining Using Chamelion\bin\Frames \36.bmp	46031493
▶		

Fig 7. Frame Value Calculation.

	id	frame	numclus	time
	1	0	3	3:51:44 PM
	2	1	2	3:51:45 PM
	3	2	2	3:51:47 PM
	4	3	3	3:51:48 PM
	5	4	2	3:51:49 PM
	6	5	2	3:51:51 PM
▶	(AutoNumber)	0		

Fig 8. Number of Cluster Formation – Graphic Video File.

	id	frame	numclus	time
	1	0	7	3:53:42 PM
	2	1	6	3:53:44 PM
	3	2	5	3:53:47 PM
	4	3	5	3:53:49 PM
	5	4	4	3:53:52 PM
	6	5	5	3:53:54 PM
	7	6	6	3:53:57 PM
	8	7	5	3:53:59 PM
	9	8	6	3:54:02 PM
	10	9	6	3:54:04 PM
	11	10	5	3:54:07 PM
	12	11	4	3:54:09 PM
	13	12	5	3:54:12 PM
	14	13	4	3:54:14 PM
▶	(AutoNumber)	0		

Fig 9. Number of Cluster Formation – News Video File.

References

1. Saravanan D, Ramesh Kumar A, "ContentBased Image Retrieval using Color Histogram", International journal of computer science and information technology . 2013 April; 4(2), 242-45.

2. Saravanan D, Dr. Srinivasan S, "Video information retrieval using :CHEMELEON Clustering" International Journal of Emerging Trends & Technology in Computer Science.2013Jan-Feb;2(1), 166-170.
3. W. Chu et.al ," An iterative image registration algorithm by optimizing similarity measurement", Journal of Research of the National Institute of Standards and Technology.2010 Jan-Feb; 115 (1), 1–6.
4. Shabana Asmi, P., Justin Samuel, S. (2015), "An analysis and accuracy prediction of heart disease with association rule and other data mining techniques ",Journal of Theoretical and Applied Information Technology, Vol.79. No.2, pp. 254-260.
5. Saravanan D Effective Multimedia Content Retrieval, International Journal of Applied Environmental Sciences.2015 Sep-Oct;10(5), 1771-17783.
6. Liujuan Cao a, RongrongJi b,n, YueGao c, WeiLiu b, QiTian, "Mining spatiotemporal video patterns towards robust action retrieval", Neurocomputing .2013April; 105(2013), 61-69.
7. Saravanan D ,Dr.Srinivasan S ,"A proposed new Algorithm for analysis for analysis of Hierarchical clustering in video Data mining" International journal of Data mining and knowledge engineering.2011 July;3(9), 569-572.
8. Saravanan D,Somasundaram V, "Matrix Based Sequential Indexing Technique for Video Data Mining "Journal of Theoretical and Applied Information Technology. 2014 Nov ,67(3), 725-31.